ELK 由三部分组成elasticsearch、logstash、kibana，一个一个的啃过去。在安装elasticsearch之前，首先来看看elasticsearch几个至关重要的术语。

**NRT**

elasticsearch是一个近似实时的搜索平台，从索引文档到可搜索有些延迟，通常为1秒。

**集群**

集群就是一个或多个节点存储数据，并提供跨节点的联合索引和搜索的功能。集群有一个唯一性标示的名字，默认是elasticsearch，集群名字很重要，每个节点是基于集群名字加入到其集群中的。因此，确保在不同环境中使用不同的集群名字。

一个集群可以只有一个节点。

强烈建议在配置elasticsearch时，配置成集群模式。

**节点**

节点就是一台单一的服务器，是集群的一部分，存储数据并参与集群的索引和搜索功能。像集群一样，节点也是通过名字来标识，默认是在节点启动时随机分配的字符名。当然啦，你可以自己定义。该名字也蛮重要的，在集群中用于识别服务器对应的节点。

节点可以通过指定集群名字来加入到集群中。默认情况下，每个节点被设置成加入到elasticsearch集群。如果启动了多个节点，假设能自动发现对方，他们将会自动组建一个名为elasticsearch的集群。

**索引**

索引是有几分相似属性的一系列文档的集合。如[nginx](http://www.ttlsa.com/nginx/)日志索引、syslog索引等等。索引是由名字标识，名字必须全部小写。这个名字用来进行索引、搜索、更新和删除文档的操作。

索引相对于关系型数据库的库。

**类型**

在一个索引中，可以定义一个或多个类型。类型是一个逻辑类别还是分区完全取决于你。通常情况下，一个类型被定义成具有一组共同字段的文档。如ttlsa运维生成时间所有的数据存入在一个单一的名为logstash-ttlsa的索引中，同时，定义了用户数据类型，帖子数据类型和评论类型。

类型相对于关系型数据库的表。

**文档**

文档是信息的基本单元，可以被索引的。文档是以JSON格式表现的。

在类型中，可以根据需求存储多个文档。

虽然一个文档在物理上位于一个索引，实际上一个文档必须在一个索引内被索引和分配一个类型。

文档相对于关系型数据库的列。

**分片和副本**

在实际情况下，索引存储的数据可能超过单个节点的硬件限制。如一个十亿文档需1TB空间可能不适合存储在单个节点的磁盘上，或者从单个节点搜索请求太慢了。为了解决这个问题，elasticsearch提供将索引分成多个分片的功能。当在创建索引时，可以定义想要分片的数量。每一个分片就是一个全功能的独立的索引，可以位于集群中任何节点上。

分片的两个最主要原因：

1. 水平分割扩展，增大存储量
2. 分布式并行跨分片操作，提高性能和吞吐量

分布式分片的机制和搜索请求的文档如何汇总完全是有elasticsearch控制的，这些对用户而言是透明的。

网络问题等等其它问题可以在任何时候不期而至，为了健壮性，强烈建议要有一个故障切换机制，无论何种故障以防止分片或者节点不可用。

为此，elasticsearch让我们将索引分片复制一份或多份，称之为分片副本或副本。

副本也有两个最主要原因：

1. 高可用性（HA），以应对分片或者节点故障。出于这个原因，分片副本要在不同的节点上。
2. 提供性能，增大吞吐量，搜索可以并行在所有副本上执行。

总之，每一个索引可以被分成多个分片。索引也可以有0个或多个副本。复制后，每个索引都有主分片(母分片)和复制分片(复制于母分片)。分片和副本数量可以在每个索引被创建时定义。索引创建后，可以在任何时候动态的更改副本数量，但是，不能改变分片数。

默认情况下，elasticsearch为每个索引分片5个主分片和1个副本，这就意味着集群至少需要2个节点。索引将会有5个主分片和5个副本(1个完整副本)，每个索引总共有10个分片。

每个elasticsearch分片是一个Lucene索引。一个单个Lucene索引有最大的文档数LUCENE-5843, 文档数限制为2147483519(MAX\_VALUE - 128)。 可通过\_cat/shards来监控分片大小。

[](http://www.ttlsa.com/wp-content/uploads/2015/11/elk2.jpg)

这里所言的复制和副本是同一个意思。

这部分内容比较枯燥无味，不过蛮重要的，贯穿于整个elasticsearch。